Ransomware 3.0 se une a la AI activa, lo que hace que la ciberresiliencia sea más crítica que nunca

Ransomware 3.0 es una nueva generación de ransomware que utiliza AI para planificar y organizar acciones de rescate: y como vimos esta semana, las mismas capacidades de agente que hacen que la AI sea valiosa para la productividad y la defensa ahora se pueden armar para la ofensa a una velocidad y escala sin precedentes.

La era 3.0 también marca el cambio de los atacantes de ransomware a los navegadores de AI, donde los tipos malos creen que atraerán a más víctimas. El nuevo giro sobre ransomware también destaca la disposición de los perpetradores de usar varias formas de extorsión en conjunto (a veces denominadas extorsión triple) para aplicar una mayor presión sobre los objetivos.

Ransomware pasa al navegador

Los navegadores de AI tienen funciones como chatbots incorporados y agentes de rendimiento de tareas. Los usuarios están exaltando las virtudes de los navegadores de AI. Pero las funciones que hacen que los navegadores de AI sean inteligentes ya están siendo modificadas por hackers de ransomware. Con la "inyección inmediata", los atacantes crean códigos para aplicaciones creadas en modelos de lenguaje grande (LLM), engañando a la AI agente para que piense que la solicitud proviene de un usuario de confianza.

Los investigadores de la Escuela de Ingeniería Tandon de la Universidad de Nueva York crearon recientemente un software de prueba de concepto de inyección rápida, que fue "descubierto" por investigadores externos y rápidamente llamado "PromptLock". (Los investigadores pensaron que habían descubierto un verdadero ransomware creado por hackers ingeniosos antes de que NYU hablara y dijera: "No, lo hicimos").

Como parte del proyecto, los investigadores de la NYU simularon las cuatro fases de los ataques de ransomware, mapear sistemas, identificar archivos valiosos, robar o encriptar datos y generar notas de rescate, con modelos de lenguaje grande (LLM). Los investigadores probaron su simulación impulsada por AI en computadoras personales, servidores empresariales y sistemas de control industrial y descubrieron que los ataques de ransomware completos podían

ejecutarse de forma autónoma.

Informe de Espionaje Orquestado por AI de Anthropic

En un eco del mundo real de esos hallazgos de investigación, esta semana Anthropic informó descubrir la primera campaña de ciberespionaje orquestada por AI a gran escala, un ataque supuestamente iniciado por un grupo patrocinado por el estado que utiliza agentes de AI para infiltrarse en las principales empresas de tecnología, bancos, fabricantes y organismos gubernamentales. A diferencia de los incidentes anteriores en los que los humanos dirigían el flujo de ataque, esta operación utilizaba un sistema de AI autónomo para realizar hasta el 90 % de la campaña por sí solo, desde el reconocimiento hasta la exfiltración de datos. Esto destaca la rapidez con la que evolucionan las amenazas cibernéticas impulsadas por AI. Para los defensores, esto marca un momento abrumador: prueba de que el panorama de amenazas no solo está cambiando, sino que está acelerando más allá de los ciclos tradicionales de respuesta humana.

Lectura relacionada: Cuando falla la privacidad del nivel de aplicación: El caso de la gobernanza de capa de datos

No toda la AI es buena AI

La AI se puede aplicar a casi todas las fases de un ataque de ransomware. El phishing impulsado por la AI puede ser más matizado, más contextualmente correcto y mejor dirigido a “enlaces débiles” u objetivos de alto valor, conocidos como phishing selectivo. Una vez que se ha obtenido acceso a una empresa, la AI no solo puede observar las actividades de los objetivos de alto valor, sino que también puede buscar formas de permitir la evasión, como imitar la actividad legítima o buscar errores de configuración que ofrezcan un pie en la puerta de aún más activos.

Además, la capacidad muy afamada de la AI para escribir código es especialmente útil para los ciberdelincuentes. Por ejemplo, pueden implementar su malware con una diferenciación aleatoria suficiente para engañar los métodos tradicionales de detección basados en firmas. Dicho código polimórfico derivado de AI puede mejorar o redirigir de forma continua el ransomware según sea necesario para la evasión o la focalización oportunista.

Por último, la AI permite a los delincuentes automatizar gran parte del trabajo de lanzar ataques de ransomware, un punto crucial cuando considera que muchos

esquemas cibernéticos aumentan sus posibilidades de éxito cuando aprovechan la escalabilidad fácil, especialmente en la etapa de phishing. La capacidad de automatizar es un multiplicador de fuerza, lo que permite que se sigan realizando más operaciones simultáneamente.

Agregar capas de extorsión al Ransomware

Otra característica del ransomware 3.0 son las formas de coerción más agresivas y específicas. Cuando el ransomware era nuevo, la primera ola de atacantes era el contenido para encriptar los datos y solicitar el pago a cambio de la clave de descifrado. En muchos casos, el cifrado era de mala calidad y podía romperse. Además, un sistema riguroso de copias de seguridad podría permitir que las organizaciones ignoren las demandas y simplemente implementen la recuperación.

En la siguiente etapa del ransomware, o ransomware 2.0, los delincuentes no solo encriptaron los datos, sino que los exfiltraron y amenazaron con liberarlos en la web pública o venderlos a uno de los muchos corredores de datos en el ecosistema de delitos cibernéticos. Ransomware como servicio también era una característica del ransomware 2.0. Con el auge de la tecnología de cadena de bloques, la metodología de encriptación dio un gran salto. En este caso, las copias de seguridad de datos podrían permitir la recuperación, pero una gran cantidad de datos que caen en manos criminales daría lugar a un enorme daño a la reputación, sin mencionar las multas regulatorias y las posibles liquidaciones financieras para los clientes y socios afectados.

Restaurar las copias de seguridad de datos en la era del ransomware 3.0 puede abollar seriamente los presupuestos empresariales. El mayor costo para las empresas que sufren un ciberataque es la recuperación de las copias de seguridad actualizadas de datos críticos, como se detalla en un nuevo estudio de resiliencia cibernética de Pure Storage y Ponemon Institute. Los encuestados dijeron que otros costos significativos son recuperar, reparar o reemplazar los sistemas y aplicaciones afectados, detectar y contener el incidente y realizar pruebas para garantizar que los sistemas restaurados funcionen correctamente.

Los peligros que representa el ransomware 3.0 son reales. Los delincuentes que se sienten emocionados por su dominio de la AI pueden volverse más agresivos al ejercer presión sobre las organizaciones objetivo.

La resiliencia cibernética es más crítica que nunca

La AI agente no solo ha reescrito las reglas para la productividad, sino que también ha reescrito los riesgos para la TI empresarial. Las empresas ahora deben lidiar con el reconocimiento automatizado, el malware adaptativo y la extorsión de varias etapas, y se está escalando más rápido de lo que cualquier equipo humano puede responder. A medida que el panorama evoluciona, los equipos de seguridad empresarial deben:

- **Reconozca el cambio de AI de acción**: Comprenda que el panorama de amenazas ahora incluye marcos de ataque totalmente autónomos, capaces de operaciones sostenidas sin supervisión humana.
- **Considere optar por no incluir funciones que permitan la memoria persistente, el acceso amplio a las herramientas o la retención de datos a largo plazo, cuando no sea crítico para el negocio.** Muchas plataformas de AI agente, incluida Anthropic, permiten que las empresas restrinjan el acceso de los agentes a datos sensibles, desactiven la memoria a largo plazo y limiten la integración con API externas o herramientas no examinadas, todo lo cual reduce directamente el riesgo si se produce un jailbreak o una inyección inmediata.
- **Fortalezca las defensas conscientes de la AI:** Invierta en herramientas de detección que puedan reconocer el comportamiento anómalo del agente, la escalada de privilegios y la automatización rápida, no solo las firmas de malware tradicionales.
- **Refuerce los sistemas contra los jailbreaks y el uso indebido:** Implemente medidas de seguridad en capas en torno al acceso a la AI, incluidas las limitaciones en el uso de herramientas, la identidad granular y los controles de roles, y la validación constante de las acciones del agente.
- **Enfóquese en la velocidad y la recuperación:** Las capacidades de respuesta rápida y recuperación instantánea son fundamentales, dado el ritmo al que los agentes autónomos pueden moverse. Las copias de seguridad inmutables y el monitoreo en tiempo real ahora son esenciales, no opciones.
- **Mantenga a los humanos informados para su supervisión:** Las herramientas autónomas requieren límites de gobernanza estrictos: los equipos de seguridad deben crear comités de supervisión, realizar revisiones periódicas y garantizar rutas de escalamiento para actividades sospechosas.
- **Comparta inteligencia y colabore:** El intercambio continuo de información con socios y organismos de la industria es vital para

contrarrestar las campañas distribuidas impulsadas por AI antes de que se propaguen o escalen.

Las empresas deben tratar este incidente como una llamada de atención. La era de la AI agente en las operaciones cibernéticas ahora es una realidad, lo que hace que la planificación de resiliencia moderna, la detección rápida y la supervisión humana sean elementos críticos para la supervivencia y el éxito.

Descargue el informe "El estado de la ciberresiliencia".

Preguntas frecuentes

¿Qué es la AI agente y por qué es relevante para la ciberseguridad empresarial?
La AI agente se refiere a sistemas autónomos dirigidos a objetivos capaces de planificar, razonar y actuar de manera independiente para ejecutar tareas complejas, incluida la detección de amenazas, la respuesta e incluso el lanzamiento de ciberataques sin supervisión humana directa. Esto permite tanto defensas de seguridad avanzadas como riesgos, ya que los atacantes ahora utilizan modelos similares para escalar las operaciones.

¿Cómo puede la AI agente aumentar la velocidad y la escala de los ataques de ransomware?
A diferencia de las herramientas de malware tradicionales, la AI agente automatiza el reconocimiento, la secuencia de comandos y la ejecución de exploits, lo que permite que las amenazas se propaguen rápidamente en las redes, lleven a cabo la triple extorsión y se adapten sobre la marcha. Las empresas pueden enfrentar ataques que evolucionan mucho más rápido de lo que pueden responder las defensas manuales.

¿Qué estrategias deben usar las empresas para defenderse contra las ciberamenazas de AI?
Las prácticas críticas incluyen monitoreo consciente de la AI en tiempo real, copias de seguridad de snapshots inmutables, controles de acceso basados en roles e intercambio proactivo de inteligencia sobre amenazas. Para obtener una guía más profunda, consulte el blog de Pure Storage sobre las mejores prácticas de ciberseguridad y adaptación.

¿Dónde encaja la infraestructura de almacenamiento en la defensa contra

los ataques de AI agente?

El almacenamiento moderno es una parte central de la resiliencia cibernética: Las snapshots de Pure Storage® SafeMode™ y las herramientas de restauración rápida ayudan a las organizaciones a recuperar datos limpios de inmediato, incluso durante ransomware o incursiones de agentes autónomos. Lea más sobre la arquitectura de almacenamiento resiliente de Pure Storage aquí.

¿Cuáles son los pasos procesables para las empresas que adoptan plataformas de AI como Anthropic?

Restrinja la herramienta del agente y el acceso a los datos, opte por dejar de usar la memoria persistente para flujos de trabajo críticos y aplique políticas estrictas de retención y revisión, minimizando la exposición si un agente es utilizado indebidamente. Los clientes de Pure Storage pueden obtener más información del último seminario web “Pregúntenos todo sobre la ciberresiliencia”.